

NHGRI GWAS catalog: description of column headings for Tab delimited File

Note: The SNP data in the catalog has been mapped to dbSNP Build 141 and Genome Assembly, GRCh38/hg38.

DATE ADDED TO CATALOG: Date added to catalog

PUBMEDID: PubMed identification number

FIRST AUTHOR: Last name of first author

DATE: Publication date (online (epub) date if available)

JOURNAL: Abbreviated journal name

LINK: PubMed URL

STUDY: Title of paper (linked to PubMed abstract)

DISEASE/TRAIT: Disease or trait examined in study

INITIAL SAMPLE SIZE: Sample size for Stage 1 of GWAS

REPLICATION SAMPLE SIZE: Sample size for subsequent replication(s)

REGION: Cytogenetic region associated with rs number (NCBI)

CHR_ID: Chromosome number associated with rs number (NCBI)

CHR_POS: Chromosomal position associated with rs number (dbSNP Build 132, NCBI)

REPORTED GENE (S): Gene(s) reported by author

MAPPED GENE(S): Gene(s) mapped to the strongest SNP (NCBI). If the SNP is located within a gene, that gene is listed. If the SNP is intergenic, the upstream and downstream genes are listed, separated by a hyphen. **UPSTREAM_GENE_ID:**

Entrez Gene ID for nearest upstream gene to rs number, if not within gene (NCBI)

DOWNSTREAM_GENE_ID: Entrez Gene ID for nearest downstream gene to rs number, if not within gene (NCBI)

SNP_GENE_IDS: Entrez Gene ID, if rs number within gene; multiple genes denotes overlapping transcripts (NCBI)

UPSTREAM_GENE_DISTANCE: distance in kb for nearest upstream gene to rs number, if not within gene (NCBI)

DOWNSTREAM_GENE_DISTANCE: distance in kb for nearest downstream gene to rs number, if not within gene (NCBI)

STRONGEST SNP-RISK ALLELE: SNP(s) most strongly associated with trait + risk allele (? for unknown risk allele). May also refer to a haplotype.

SNPS: Strongest SNP; if a haplotype is reported above, may include more than one rs number (multiple SNPs comprising the haplotype)

MERGED: denotes whether the SNP has been merged into a subsequent rs record (0 = no; 1 = yes; NCBI)

SNP_ID_CURRENT: current rs number (will differ from strongest SNP when merged = 1)

CONTEXT: SNP functional class (NCBI)

INTERGENIC: denotes whether SNP is in intergenic region (0 = no; 1 = yes; NCBI)

RISK ALLELE FREQUENCY: Reported risk allele frequency associated with strongest SNP

P-VALUE: Reported p-value for strongest SNP risk allele (linked to dbGaP Association Browser)

PVALUE_MLOG: $-\log(\text{p-value})$

P-VALUE (TEXT): Information describing context of p-value (e.g. females, smokers).

Note that p-values are rounded to 1 significant digit (for example, a published p-value of 4.8×10^{-7} is rounded to 5×10^{-7}).

OR or BETA: Reported odds ratio or beta-coefficient associated with strongest SNP risk allele

95% CI (TEXT): Reported 95% confidence interval associated with strongest SNP risk allele

PLATFORM (SNPS PASSING QC): Genotyping platform manufacturer used in Stage 1; also includes notation of pooled DNA study design or imputation of SNPs, where applicable

CNV: Study of copy number variation (yes/no)

Updated: December 4, 2014